Probabilidade e Estatística Matemática

Ano 2022 - Módulo 4 - Aula 1 (Parte 1/2)

Paulo Justiniano Ribeiro Jr

Programa de Pós Graduação
Métodos Numéricos em Engenharia
Universidade Federal do Paraná

10 de agosto de 2022

Exemplo: O problema dos testes de diagnóstico

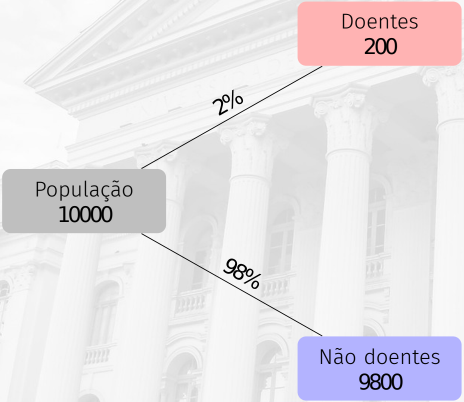
Informação disponível:

- Teste de varredura (*screening*) para uma determinada doença
- Testes são *imperfeitos*, suponha que:
  acerta 90% dos que tem doença e 80% dos que não tem.
- A doença ocorre em 2% da população

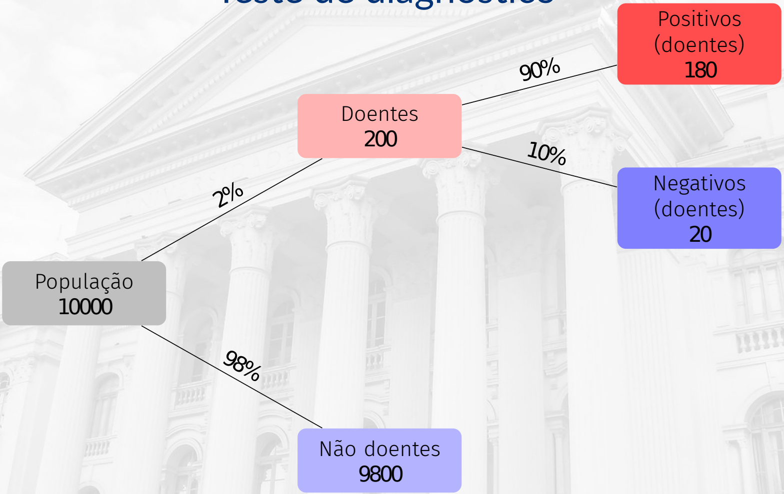
Pergunta de interesse:
Se uma pessoa testou positivo, qual a chance de ter a doença?

DEST Departamento de Estatística

Teste de diagnóstico



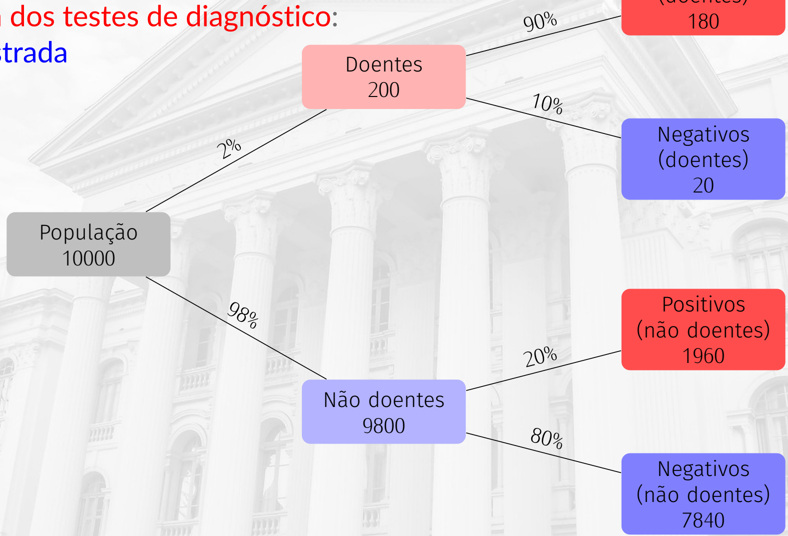
Teste de diagnóstico



Qual a probabilidade?

O problema dos testes de diagnóstico:

Solução ilustrada



Qual a probabilidade?

Terminologia específica:

- Teste de *screening* para uma determinada doença
- Teste *imperfeito*:
  acerta 90% dos que tem doença (*sensibilidade*) e portanto 10% de *falso negativo*
  acerta 80% dos que não tem (*especifidade*) e portanto 20% de *falso positivo*
- A doença ocorre em 2% da população (*prevalência*)

Pergunta de interesse:
Se uma pessoa testou positivo, qual a chance de ter a doença?
(*valor preditivo positivo*)

Testes de diagnóstico - Organizando dados tabelas

Características do teste (para diagnósticos conhecidos)

	Positivo	Negativo	
c/ Doença	0,90	0,10	1
s/ Doença	0,20	0,80	1

Temos 2% com a doença na população (e portanto 98% sem)

	Positivo	Negativo	Total
c/ Doença	0,018	0,002	0,02
s/ Doença	0,196	0,784	0,98
Total	0,214	0,786	1

Conhecendo o resultado do teste

	Positivo	Negativo
c/ Doença	0.0841	0.00254
s/ Doença	0.916	0.997
Total	1	1

DEST Departamento de Estatística

Testes de diagnóstico

A notação é nossa amiga!

$$P[+|D] = 0,90 \longrightarrow P[-|D] = 0,10$$
$$P[-|\overline{D}] = 0,80 \longrightarrow P[+|\overline{D}] = 0,20$$
$$P[D] = 0,02$$
$$P[D|+] = ?$$

E um Teorema resolve o problema!

$$\begin{aligned} P[D|+] &= \frac{P[+|D] \cdot P[D]}{P[+|D] \cdot P[D] + P[+|\overline{D}] \cdot P[\overline{D}]} \\ &= \frac{0.90 \cdot 0.02}{0.90 \cdot 0.02 + 0.20 \cdot 0.98} = 0.0841 \end{aligned}$$

DEST Departamento de Estatística

Testes de diagnóstico

Solução ilustrada (com notação)

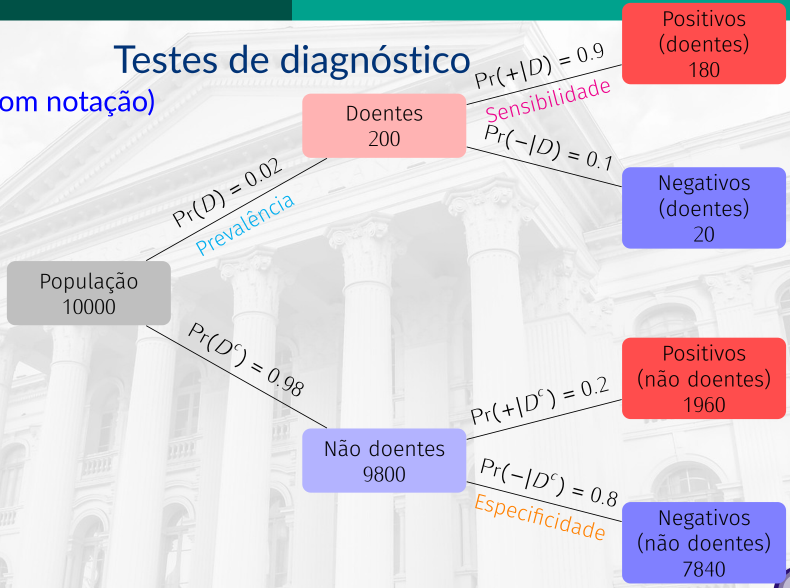


Figura 2. Ilustração do problema do teste de diagnóstico

Atualização de informação

- Existência de uma probabilidade inicial (*a priori*).

Estado	Doente (D)	Sadio ($\overline{D}$)
Probabilidade	0,02	0,98

- Informação
  Teste positivo e $P[+|D] = 0,90$ e $P[+|\overline{D}] = 0,20$
- (possivelmente) alterando probabilidades (*a posteriori*).

Estado	Doente (D)	Sadio ($\overline{D}$)
Probabilidade	0.0841	0.916

- Classificação de possíveis *estados da natureza* e suas probabilidades.
- Decisão guiada pela probabilidade.

DEST Departamento de Estatística

Adaptando notação

- Existência de uma probabilidade inicial (*a priori*).

Estado	$\theta = 1$ (Doente)	$\theta = 0$ (Sadio)
Probabilidade	0,02	0,98

- Dado/Informação $Y = 1$ se positivo, $Y = 0$ se negativo,
  Teste positivo e $P[Y = 1|\theta = 1] = 0,90$ e $P[Y = 1|\theta = 0] = 0,20$
- atualizando probabilidades (*a posteriori*).

Estado	$\theta = 1$ (Doente)	$\theta = 0$ (Sadio)
Probabilidade	0.0841	0.916

- Distribuição de probabilidades dos *valores do parâmetro* e suas probabilidades.

Teorema de Bayes

Reescrevendo e reinterpretando como problema de classificação

$$P[A_j|B] = \frac{P[B|A_j] \cdot P[A_j]}{\sum_j P[B|A_j] \cdot P[A_j]} \propto P[B|A_j] \cdot P[A_j]$$

Como o paciente deve ser classificado após o teste?

$$P[A_1|B] \propto P[B|A_1] \cdot P[A_1] \text{ (ou, } P[D|+] \propto P[+|D] \cdot P[D])$$
$$P[A_2|B] \propto P[B|A_2] \cdot P[A_2] \text{ (ou, } P[\overline{D}|+] \propto P[+|\overline{D}] \cdot P[\overline{D}])$$

Portanto

$$P[A_1|B] \propto 0,9 \cdot 0,02 = 0,018$$
$$P[A_2|B] \propto 0,2 \cdot 0,98 = 0,196$$

A_1 e A_2 são todas as categorias possíveis, as probabilidades devem somar 1:

$$P[A_1|B] = \frac{0,018}{0,018 + 0,196} = 0,084$$
$$P[A_2|B] = \frac{0,196}{0,018 + 0,196} = 0,916$$

Teorema de Bayes

E se o teste for repetido?
Notaçao: B_1 positivo no primeiro teste e B_2 positivo no segundo teste
Supondo independência

$$P[A_1|B_1, B_2] \propto P[B_2|A_1] \cdot P[B_1|A_1] \cdot P[A_1]$$
$$P[A_2|B_1, B_2] \propto P[B_2|A_2] \cdot P[B_1|A_2] \cdot P[A_2]$$

Portanto se for o mesmo teste (mesmas características de sensibilidade e especificidade)

$$P[A_1|B] \propto 0,90^2 \cdot 0,02 = 0,0162$$
$$P[A_2|B] \propto 0,20^2 \cdot 0,98 = 0,0392$$

Logo,

$$P[A_1|B] = \frac{0,0162}{0,0162 + 0,0392} = 0,292 \quad P[A_2|B] = \frac{0,0392}{0,0162 + 0,0392} = 0,708$$

A classificação ainda é a mesma mas as chances mudaram!
Com três testes positivos $P[A_1|B] = 0.65$.

Repetindo o teste - resumo

Estados da natureza:

Estado (θ)	$\theta = 0(\bar{D})$	$\theta = 1(D)$
Probabilidade	0,98	0,02

Estados da natureza após primeiro exame positivo:

Estado ($\theta \mid y_1$)	$\theta = 0(\bar{D})$	$\theta = 1(D)$
Probabilidade	0,916	0,084

Estados da natureza após segundo exame positivo:

Estado ($\theta \mid y_1, y_2$)	$\theta = 0(\bar{D})$	$\theta = 1(D)$
Probabilidade	0,708	0,292

DEST
Departamento de Estatística

Teorema de Bayes

Testes positivos e negativos?
Notaçao: B_1 positivo no primeiro teste, e B_2 negativo no segundo teste e B_3 positivo no terceiro teste.
Supondo independência:

$$P[A_1|B_1, \overline{B}_2, B_3] \propto P[B_3|A_1] \cdot P[\overline{B}_2|A_1] \cdot P[B_1|A_1] \cdot P[A_1]$$
$$P[A_2|B_1, \overline{B}_2, B_3] \propto P[B_3|A_2] \cdot P[\overline{B}_2|A_2] \cdot P[B_1|A_2] \cdot P[A_2]$$

Portanto se for o mesmo teste (mesmas características de sensibilidade e especificidade),

$$P[A_1|B] \propto 0,90 \cdot 0,10 \cdot 0.90 \cdot 0,02 = 0.00162$$
$$P[A_2|B] \propto 0,20 \cdot 0,80 \cdot 0,20 \cdot 0,98 = 0.0314$$

Logo,

$$P[A_1|B] = \frac{0.00162}{0.00162 + 0.0314} = 0.0491 \quad P[A_2|B] = \frac{0.0314}{0.00162 + 0.0314} = 0.951$$

DEST Departamento de Estatística

O problema dos testes de diagnóstico

E se o teste tivesse sido feito por recomendação médica após um exame? Mudaria algo?
Baseado em sua experiência o médico estima que 30% dos pacientes com os sintomas apresentados tem a doença.
Solução? (reproduzir passos acima!)
Neste caso:

$$
\begin{aligned}
P[D|+] &= \frac{P[+|D] \cdot P[D]}{P[+|D] \cdot P[D] + P[+|\overline{D}] \cdot P[\overline{D}]} \\
&= \frac{0,90 \cdot 0,30}{0,90 \cdot 0,30 + 0,20 \cdot 0,70} = 0.659
\end{aligned}
$$

Comparação e *interpretação* (e …muito cuidado!)

Atualização de informação

- Existência de uma probabilidade inicial (*a priori*).

Estado	$\theta = 1$ (Doente)	$\theta = 0$ (Sadio)
Probabilidade	0,30	0,70

- Informação
  Teste positivo e $P[Y = 1|\theta = 1] = 0,90$ e $P[Y = 1|\theta = 0] = 0,20$

- (possivelmente) alterando probabilidades (*a posteriori*).

Estado	$\theta = 1$ (Doente)	$\theta = 0$ (Sadio)
Probabilidade	0.659	0.341

Teorema de Bayes

$$P[A_j|B] = \frac{P[B|A_j] \cdot P[A_j]}{\sum_j P[B|A_j] \cdot P[A_j]}$$

- No exemplo só haviam duas categorias:

$$A_1(D) : \text{com a doença}, A_2(\overline{D}) : \text{sem a doença}.$$

- O resultado é mais geral, válido para várias categorias.
- Categorias são possíveis estados do sistema:
  - estados do sistema podem ser categóricos (ou ainda discretos ou enumeráveis),
  - estados do sistema podem ser em uma escala contínua.

Teorema de Bayes

$$P[\theta_j|Y] = \frac{P[Y|\theta_j] \cdot P[\theta_j]}{\sum_j P[Y|\theta_j] \cdot P[\theta_j]}$$

- No exemplo só haviam duas categorias:

$$\theta = 1(D) : \text{com a doença}, \theta = 0(\overline{D}) : \text{sadio.}$$

- O resultado é mais geral, válido para várias categorias ($j = 2, 3, 4, \ldots$).
- Categorias são possíveis valores do parâmetro:
  - valores do parâmetro podem ser categóricos (ou ainda discretos ou enumeráveis),
  - valores do parâmetro podem ser em uma escala contínua:

$$f(\theta|y) = \frac{f(y|\theta) \cdot f(\theta)}{\int f(y|\theta) \cdot f(\theta) \mathrm{d}\theta}.$$

DEST Departamento de Estatística

Exemplo: estimando uma proporção

Em uma população (considerada *infinita*) uma proporção θ de indivíduos apresenta determinada característica.
Deseja-se (inferências):

- estimar θ,
- expressar a incerteza sobre esta estimativa,
- verificar se θ (e portanto a população) está fora de normas/referências (proporção max. de 20%), se há evidências de um desvio "relevante" (significativo).

Dados de *uma* amostra (considerada aleatória):
$n = 80$ e $y = 19$
Como proceder?

Objetivos de inferência

- *Problema*, *decisão* e *evidência*.
  - Em uma população (considerada *infinita*) uma proporção θ de indivíduos apresenta determinada característica. Deseja-se saber se θ ultrapassa 0,20 (20%).
  - Dados de *uma* amostra (considerada aleatória): $n = 80$ e $y = 19$.
- Deseja-se (inferências).
  - Estimar θ, obtendo da amostra uma *estimativa* $\hat{\theta}$,
  - Expressar a incerteza sobre esta estimativa,
  - Avaliar se θ (e portanto "o estado" da população) está fora do valor referência (proporção max. de 20%), se há evidências de um desvio "relevante" (*significativo*).
- Como proceder?
  - Princípios gerais (paradigma(s)) a serem seguidos?

DEST Departamento de Estatística

Aprendizado estatístico

Questões:

- O que os dados dizem?
- Em que devo acreditar?
- O que devo fazer?

Objetivos:

Estimativa de θ,
expressão da incerteza,
opinião em relação a valor de interesse $\theta_0 = 0,20$

DEST Departamento de Estatística

O que os dados dizem?

Se a proporção é θ, podemos avaliar a chance (probabilidade) de obter um certo número Y de indivíduos com a característica em uma amostra de n indivíduos. Sob certas suposições é razoável adotar:

$$P[Y = y|\theta] = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$$

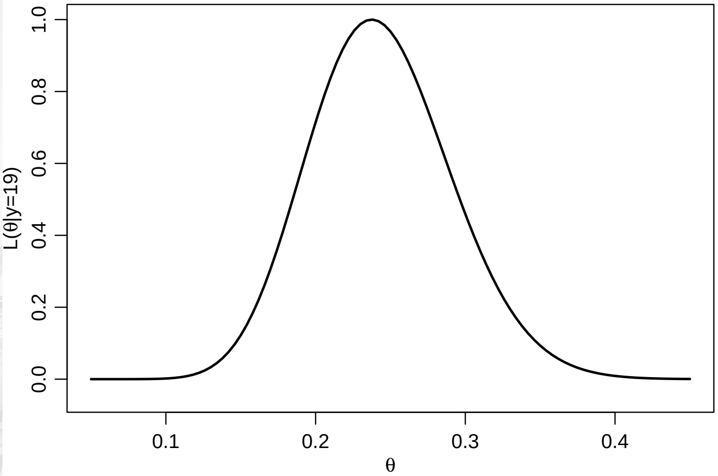
Quando obtemos a amostra temos $n = 80$ e $y = 19$. Para cada θ temos então a chance de obter esta amostra.

$$P[Y = y|\theta] = \binom{80}{19}\theta^{19}(1-\theta)^{80-19}$$

Como esta chance muda para cada valor de θ?
Considerando todos os θ possíveis temos uma função de θ:

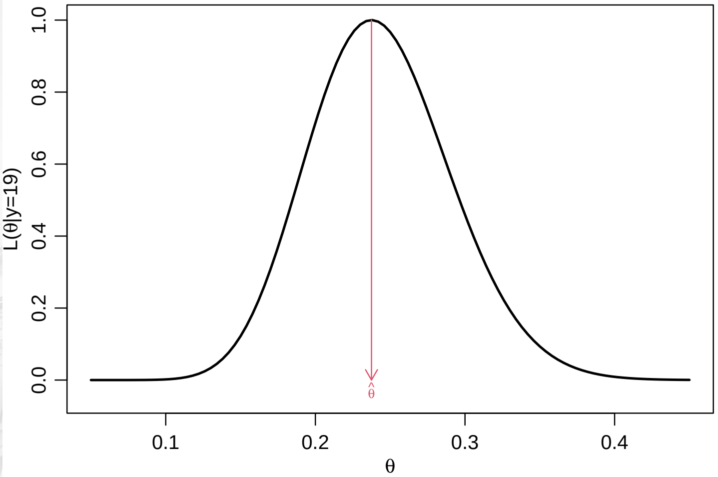
$$L[\theta|y] = \binom{80}{19}\theta^{19}(1-\theta)^{80-19}$$

Uma função representando ideias



Objetivos:
$\rightarrow$ Estimativa de θ,
$\rightarrow$ expressão da incerteza,
$\rightarrow$ opinião em relação a valor de interesse $\theta_0 = 0,20$

Uma função representando ideias



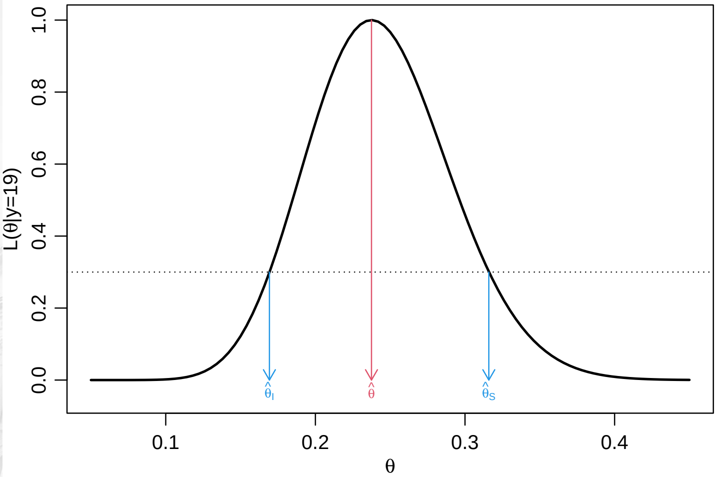
Objetivos:

$\rightarrow$ **Estimativa de** $\theta \rightarrow$ expressão da incerteza,

$\rightarrow$ opinião em relação a valor de interesse

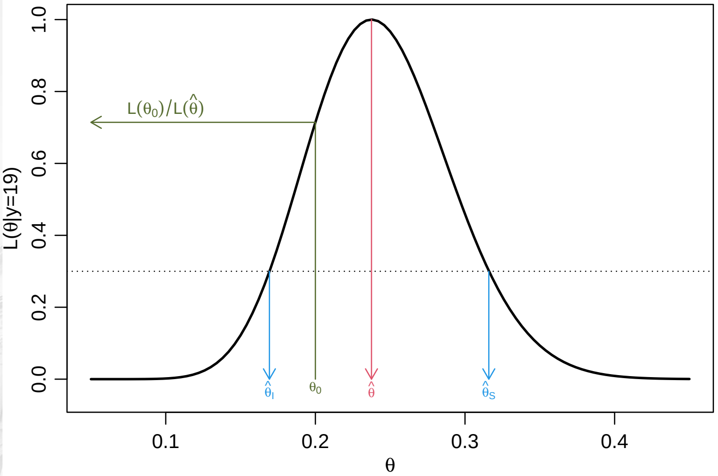
$\theta_0 = 0,20$

Uma função representando ideias



Objetivos:
$\rightarrow$ Estimativa de θ,
$\rightarrow$ **expressão da incerteza**

Uma função representando ideias



Objetivos:
$\rightarrow$ Estimativa de θ,
$\rightarrow$ expressão da incerteza,
$\rightarrow$ **opinião em relação a valor de interesse**
$\theta_0 = 0,20$

Suposições e considerações adicionais

- Suposições
  - amostra aleatória (independência)
  - população *infinita*
  - modelo binomial (critério de parada: n fixo) e invariância
- Critérios necessários
  - onde efetuar o corte da função para determinar faixas de incerteza?
  - como avaliar o valor de verossimilhança (relativa) para θ_0?

Suposições e considerações adicionais

- Suposições:
  - amostra aleatória (independência),
  - população *infinita*,
  - modelo binomial (critério de parada: n fixo) e invariância.
- Critérios necessários:
  - onde efetuar o corte da função para determinar faixas de incerteza,
  - como avaliar o valor de verossimilhança (relativa) para θ_0,
- Indo além dos dados.
  - Não há ou não se usa nenhuma informação "acessória/preliminar" sobre θ?
  - Como se comportariam outras amostras que fossem eventualmente tomadas?
  - Questões motivam e caracterizam diferentes abordagens!

Inferência Bayesiana

O objeto de inferência é a distribuição à posteriori

- A incerteza inicial sobre θ é expressa na forma de uma distribuição priori para θ
- Com amostra atualizamos opinião θ com a informação contida na verossimilhança
- O conhecimento/incerteza atualizados sobre θ é expresso pela distribuição posteriori

Formalmente:

$$f(\theta|y) \propto f(\theta) \cdot L(\theta|y)$$

ou, usando jargão técnico:

$$\text{posteriori} \propto \text{priori} \cdot \text{verossimilhança}$$

A essência de Bayes ilustrada (I)

Exemplo I : estimação da proporção de atributo (θ) na população

- **Priori**: Acredita-se que o atributo ocorre em 40% da população com 70% de chance de estar entre 30 e 50%.
  Informação expressa como distribuição de probabilidades para θ:

$$[\theta] \sim \mathrm{Beta}(10; 15) \longrightarrow f(\theta) = \mathrm{C}\ \theta^{10-1}(1-\theta)^{15-1}$$

- **Verossimilhança**: Modelo Binomial, amostra n=80 e y=19

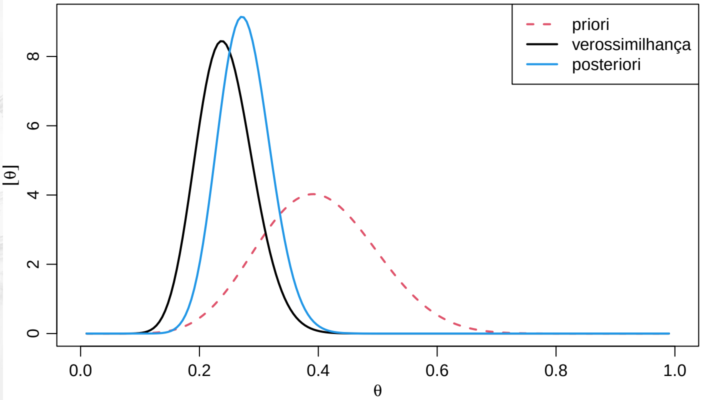
$$L[\theta] = \binom{80}{19}\theta^{19}(1-\theta)^{80-19}$$

- **Posteriori**: a distribuição de probabilidades para θ após observar os dados:

$$[\theta|y] \sim \mathrm{Beta}(29; 76) \longrightarrow f(\theta|y) = \mathrm{C}\ \theta^{29-1}(1-\theta)^{76-1}$$

C é um valor constante conhecido.

A essência de Bayes ilustrada (I)



A essência de Bayes ilustrada (I)

```r
(prI <- prioriBeta(0.4, c(0.30, 0.50), 0.70))
##     alpha      beta
##  9.912277 14.868415
postBinom(19, 80, prI, plot=FALSE)
## $pars
##                alpha     beta
## priori      9.912277 14.86842
## posteriori 28.912277 75.86842
##
## $summary
##                 moda     media   variancia
## priori     0.3912206 0.4000000 0.009309292
## posteriori 0.2715712 0.2759313 0.001888750
##
## $EMV
## [1] 0.2375
```

A essência de Bayes ilustrada (II)

Uma priori bem diferente:

- **Priori**: Acredita-se que o atributo ocorre em 8% da população com 90% de chance de estar entre 3 e 20%.

$$[\theta] \sim \mathrm{Beta}(2; 24) \longrightarrow f(\theta) = C\ \theta^{2-1}(1-\theta)^{24-1}$$

- **Verossimilhança**: Modelo Binomial, amostra n=80 e y=19

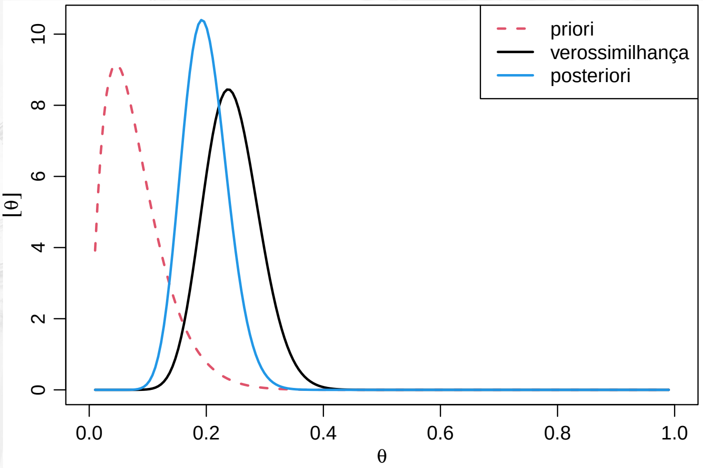
$$L[\theta] = \binom{80}{19}\theta^{19}(1-\theta)^{80-19}$$

- **Posteriori**: após observar os dados:

$$[\theta|y] \sim \mathrm{Beta}(21; 85) \longrightarrow f(\theta|y) = C\ \theta^{21-1}(1-\theta)^{85-1}$$

C é um valor constante conhecido.

A essência de Bayes ilustrada (II)



A essência de Bayes ilustrada (II)

```
(prII <- prioriBeta(0.08, c(0.03, 0.20), 0.90))
##     alpha      beta
##  2.124901 24.436358
postBinom(19, 80, prII, plot=FALSE)
## $pars
##                alpha     beta
## priori      2.124901 24.43636
## posteriori 21.124901 85.43636
##
## $summary
##                 moda     media   variancia
## priori     0.0457998 0.0800000 0.002670415
## posteriori 0.1924700 0.1982418 0.001477688
##
## $EMV
## [1] 0.2375
```

A essência de Bayes ilustrada (III)

Uma priori vaga :

- **Priori**: Não se sabe praticamente nada sobre θ. Expressa-se então que o atributo ocorre em 50% da população mas com 90% de chance de estar entre 5 e 95%.

$$[\theta] \sim \mathrm{Beta}(1.2, 1.2) \longrightarrow f(\theta) = \mathrm{C}\ \theta^{1.2-1}(1-\theta)^{1.2-1}$$

- **Verossimilhança**: Modelo Binomial, amostra n=80 e y=19

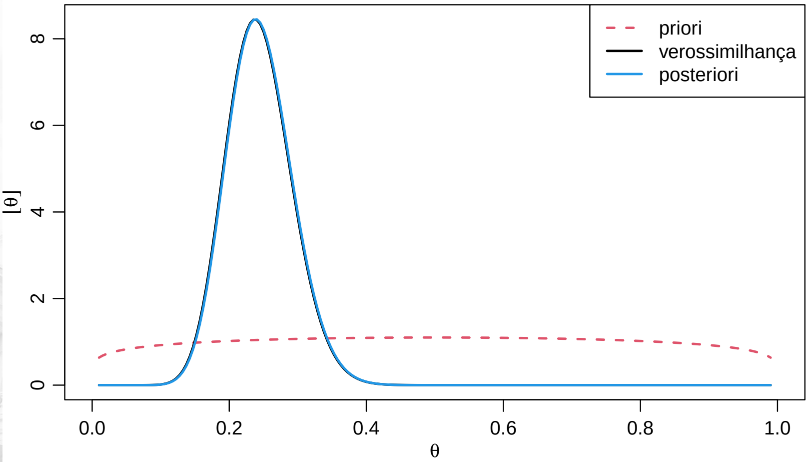
$$L[\theta] = \binom{80}{19} \theta^{19}(1-\theta)^{80-19}$$

- **Posteriori**: após observar os dados:

$$[\theta|y] \sim \mathrm{Beta}(20, 62) \longrightarrow f(\theta|y) = \mathrm{C}\ \theta^{20-1}(1-\theta)^{62-1}$$

C é um valor constante conhecido.

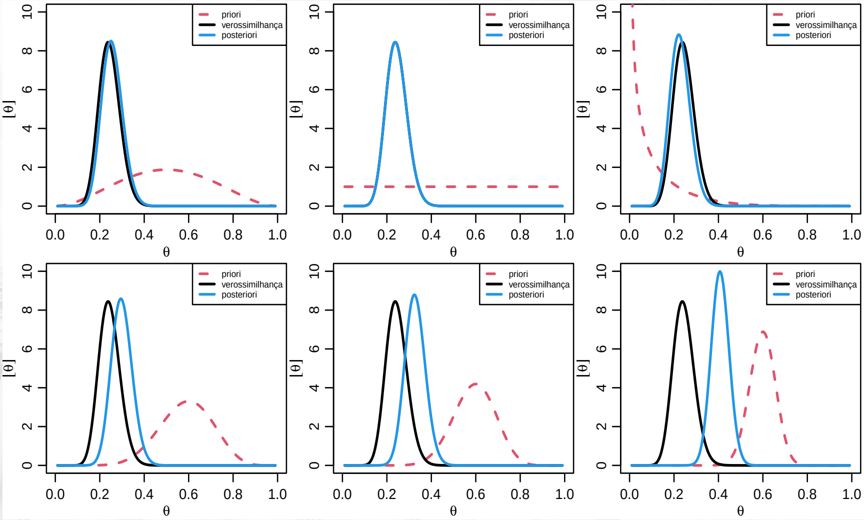
A essência de Bayes ilustrada (III)



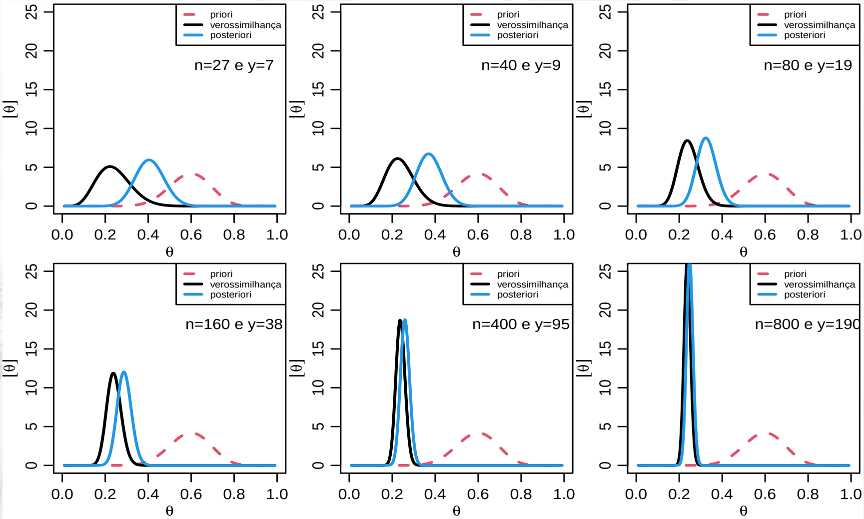
A essência de Bayes ilustrada (III)

```
(prIII <- prioriBeta(0.50, c(0.05, 0.95), 0.90))
##    alpha     beta
## 1.170088 1.170088
postBinom(19, 80, prIII, plot=FALSE)
## $pars
##                alpha      beta
## priori      1.170088  1.170088
## posteriori 20.170088 62.170088
##
## $summary
##                 moda     media   variancia
## priori     0.5000000 0.5000000 0.074846333
## posteriori 0.2386115 0.2449605 0.002219276
##
## $EMV
## [1] 0.2375
```

Efeito da priori (fixando amostra)



Efeito do tamanho da amostra (fixando priori)



Comentários

- Expressão da opinião "a priori" é necessária e sua especificação é um desafio.
- As interpretações de intervalo de confiança são agora probabilísticas, por exemplo pode-se avaliar intervalo (de credibilidade) para uma certa probabilidade:
$$P[a < \theta < b] = 0.95.$$
- No contexto do exemplo, pode-se avaliar
$$P[\theta \geq 0, 20].$$

DEST Departamento de Estatística

Inferência Bayesiana

Em resumo:

- **Estimativa de θ**: alguma medida resumo da posteriori (média, moda, mediana, ...)
- **expressão da incerteza**: variabilidade da distribuição posteriori
- **opinião em relação a valor de interesse $\theta \geq 0.20$**: probabilidade na posteriori